APPLICATION UNDER UNITED STATES PATENT LAWS

Invention: **METHOD AND APPARATUS FOR REGULARIZING MEASURED HRTF FOR SMOOTH 3D DIGITAL AUDIO**

Inventor: Jiashu CHEN

Farkas & Manelli P.L.L.C.
2000 M Street, N.W.
7th Floor
Washington, D.C. 20036-3307
Attorneys
Telephone: (202) 261-1000

This is a:

[ ] Provisional Application

[X] Regular Utility Application

[ ] Continuing Application

[ ] PCT National Phase Application

[ ] Design Application

[ ] Reissue Application

[ ] Plant Application

MARKED-UP COPY OF SPECIFICATION

METHOD AND APPARATUS FOR REGULARIZING MEASURED HRTF FOR SMOOTH 3D DIGITAL AUDIO

This application ~~claims priority from~~is a continuation of U.S. Patent Application No. ~~60/065,855 entitled "Multipurpose Digital Signal Processing System"~~09/191,179 entitled "Method and Apparatus for Regular Rising Measured HTRF for Smooth 3D Digital Audio" filed November 14, ~~1997,~~1998, the specification of which is explicitly incorporated herein by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates generally to three dimensional (3D) sound. More particularly, it relates to an improved regularizing model for head-related transfer functions (HRTFs) for use with 3D digital sound applications.

2. Background of Related Art

~~Many high end~~Some newly emerging consumer audio devices provide the option for three-dimensional (3D) sound, allowing a more realistic experience when listening to sound. In some applications, 3D sound allows a listener to perceive motion of an object from the sound played back on a 3D audio system.

~~Atal and Schroeder established cross-talk canceler technology as early as 1962, as described in U.S. Patent No. 3,236,949, which is explicitly incorporated herein by reference. The Atal Schroeder 3D sound cross-talk canceler was an analog implementation using specialized analog amplifiers and analog filters. To gain better sound positioning performance using two loudspeakers, Atal and Schroeder included empirically determined frequency dependent filters. Without~~

~~doubt, these sophisticated analog devices are not applicable for use with today's digital audio technology.~~

~~Interaural time difference (ITD), i.e., the difference in time that it takes for a sound wave to reach both ears, is an important and dominant parameter used in 3D sound design. The interaural time difference is responsible for introducing binaural disparities in 3D audio or acoustical displays. In particular, when a sound object moves in a horizontal plane, a continuous interaural time delay occurs between the instant that the sound object impinges upon one of the ears and the instant that the same sound object impinges upon the other ear. This ITD is used to create aural images of sound moving in any desired direction with respect to the listener.~~

~~The ears of a listener can be 'tricked' into believing sound is emanating from a phantom location with respect to the listener by appropriately delaying the sound wave with respect to at least one ear. This typically requires appropriate cancellation of the original sound wave with respect to the other ear, and appropriate cancellation of the synthesized sound wave to the first ear.~~

~~A second parameter in the creation of 3D sound is adaptation of the 3D sound to the particular environment using the external ear's free field to eardrum transfer functions, or what are called head-related transfer functions (HRTFs). HRTFs relate to the modeling of the particular environment of the user, including the size and orientation of the listeners head and body, as they affect reception of the 3D sound. For instance, the size of a listener's head, their torso, what they wear, etc., forms a form of filtering which can change the effect of the 3D sound on the particular user. An appropriate HRTF adjusts for the particular environment to allow the best 3D sound imaging possible.~~

~~The HRTFs are different for each location of the source of the sound. Thus, the magnitude and phase spectra of measured HRTFs~~

~~vary as a function of sound source location. Hence, it is commonly acknowledged that the HRTF introduces important cues in spatial hearing.~~

~~Advances in computer and digital signal processing technology have enabled researchers to synthesize directional stimuli using HRTFs. The HRTFs can be measured empirically at thousands of locations in a sphere surrounding the 3D sound environment, but this proves to require an excessive amount of processing. Moreover, the number of measurements can be very large if the entire auditory space is to be represented on a fine grid. Nevertheless, measured HRTFs represent discrete locations in a continuous auditory space.~~ Extensive research has established that human localize sound source location by using three major acoustic cues, the interaural time difference (ITD), interaural intensity difference (IID), and head-related transfer functions (HRTFs). Note that the time domain equivalent of HRTF is usually termed head-related impulse response (HRIR). Both HRTF and HRIR are interchangeably used in this invention wherever they fit the context. These cues, in turn, are used in generating 3D sound in 3D audio systems. Among these three cues, ITD and IID occur when sound, from a source in space, arrive at both ears of a listener. When the source is at a arbitrary location in space, the sound wave arrives at both ears with different time delays due the unequal path length of wave propagation. This creates the ITD. Also, due to the head shadowing effects, the intensity of the sound waves arriving at both ears can be unequal. This creates the IID.

When the sound source is in the median plane of the head, both ITD and IID become trivial. However, the listener still can localize sound in terms of its elevation, and some degree of lateralization. This effect, confirmed by recent research, is due to the filtering effects of head, torso, shoulders, and more importantly, the pinnae, collectively termed as external ear. In particular, external ear can be viewed as a set of

acoustical resonators, the resonance frequency of each equivalent resonator varies with respect to the in-coming angle of the sound source. Verified by measured HRTFs, these resonance frequencies manifest themselves as peaks and valleys in the spectra of the measured HRTFs. Moreover, these peaks and valleys change their center frequency with respect to sound source position change.

In order to synthesize a positioned 3D audio source, a particular set of ITD, IID, and a pair of HRTF has to be used. In order to simulate the motion of the sound source, in addition to the varying ITD and IID, many HRTF pairs have to be used to obtain a continuous moving sound image. In the prior arts, hundreds or thousands of measured HRTFs are used to fulfill this purpose. There are problems with this approach. This first problem is that the HRTFs are obtained with sound source at discrete locations in the space, thus not providing continuum of the HRTF function. The second problem is that the measured HRTFs contain measurement error and thus are not smooth. Both problems cause annoying clicks in simulating sound source motion, when discontinued HRTFs are switched in and out of the filtering loop.

One conventional solution to the adaptation of a discretely measured HRTF within a continuous auditory space is to "interpolate" the measured HRTFs by linearly weighting the neighboring impulse responses. This can provide a small step size for incremental changes in the HRTF from location to location. However, interpolation is conceptually incorrect because it does not account for ~~environmental changes between measured points, and thus may not provide a suitable 3D sound rendering.~~

the fact that linear combination of adjacent impulse responses increases the number of overall peaks and valleys involved, and thus significantly compromises the quality of the interpolated HRTF. This method, called direct convolution, is shown in Fig. 3. In particular, 460 is the sound

source to be 3D positioned. 410 and 412 are left channel and right channel delays, together to form ITD. 420 and 422 are left and right ear HRTFs. 430 and 432 are signals either can be sent to left and right ear for listening or can be sent to next stage for further processing.

Other attempted solutions include using one HRTF for a large area of the three-dimensional space to reduce the frequency of discontinuities which may cause a clicking sound. However, again, such solutions compromise the overall quality of the 3D sound rendering.

~~Another solution wherein spatial characteristic functions are combined directly with Eigen functions to provide a set of HRTFs is shown in Fig. 3.~~

~~In particular, a set N of Eigen filters **422-426** are combined with corresponding sets of spatial characteristic function (SCF) samples **412-416** and summed in a summer **440** to provide an HRTF (or HRIR) filter **450** which acts on a sound source **460**. The desired location of a sound image is controlled by varying the sound source elevation and/or azimuth in the sets of SCF samples **412-416**. Unfortunately, this technique is susceptible to discontinuities in the continuous auditory space as well.~~

There is thus a need for a more accurate HRTF model which provides a suitable HRTF for source locations in a continuous auditory space, without annoying discontinuities.

SUMMARY OF THE INVENTION

In accordance with the principles of the present invention, a head-related transfer function or head-related impulse response model for use with 3D sound applications comprises a plurality of ~~Eigen filters.~~eigen filters EFs). A plurality of spatial characteristic functions (SCFs) are adapted to be respectively combined with the plurality of Eigen filters. A plurality of regularizing models are adapted to regularize the plurality of

spatial characteristic functions prior to the respective combination with the plurality of Eigen filters.

A method of determining ~~spatial characteristic sets~~SCFs for use in a head-related transfer function model or a head-related impulse response model in accordance with another aspect of the present invention comprises constructing a covariance data matrix of a plurality of measured~~head-~~ head-related transfer functions or a plurality of measured head-related impulse responses. An Eigen decomposition of the covariance data matrix is performed to provide a plurality of ~~Eigen vectors.~~eigen filters. At least one principal Eigen vector is determined from the plurality of ~~Eigen vectors.~~eigen filters. The measured head-related transfer functions or head-related impulse responses are projected to the at least one principal Eigen ~~vector~~filter to create the spatial characteristic ~~sets.~~

sets. The SCF sample sets are fed into a generalized spline model for regularization for interpolation and smoothing. The regularized SCFs are then linearly combined with EFs to generate HRTFs or HRIRs that both continuous and smooth for a high quality and click-free 3D audio rendering.

BRIEF DESCRIPTION OF THE DRAWINGS

Features and advantages of the present invention will become apparent to those skilled in the art from the following description with reference to the drawings, in which:

Fig. 1 shows an implementation of a plurality of Eigen filters to a plurality of regularizing models each based on a set of SCF samples, to provide an HRTF model having varying degrees of smoothness and generalization, in accordance with the principles of the present invention.

Fig. 2 shows a process for determining the principle Eigen vectors to provide Eigen filters used in the Eigen filters shown in Fig. 1, in accordance with the principles of the present invention.

Fig. 3 shows a conventional solution wherein ~~spatial characteristic functions are combined directly with Eigen functions to provide a set of HRTFs.~~direct convolution of dry signal and HRTFs to provide 3D positioned audio signals.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Conventionally measured ~~HRTFs~~HRIRs are obtained by presenting a stimulus through a loudspeaker positioned at many locations in a three-dimensional space, and at the same time collecting responses from a microphone embedded in a mannequin head or a real human subject. To simulate a moving sound, a continuous ~~HRTF~~HRIR that varies with respect to the source location is needed. However, in practice, only a limited number of ~~HRTFs~~HRIRs can be collected in discrete locations in any given 3D space.

Limitations in the use of measured ~~HRTFs~~HRIRs at discrete locations have led to the development of functional representations of the ~~HRTFs,~~HRIRs, i.e., a mathematical model or equation which represents the ~~HRTF~~HRIR as a function of ~~frequency~~time and direction. Simulation of 3D sound is then performed by using the model or equation to obtain the desired HRIR or HRTF.

Moreover, when discretely measured ~~HRTFs~~HRIRs are used, annoying discontinuities can be perceived by the listener from a simulated moving sound source as a series of clicks as the sound object moves with respect to the listener. Further analyses indicates that the discontinuities may be the consequence of, e.g., instrumentation error, under-sampling of the three-dimensional space, a non-individualized head model, and/or a processing error. The present invention provides an

improved ~~HRTF~~HRIR modeling method and apparatus by regularizing the spatial attributes extracted from the measured ~~HRTFs~~HRIRs to obtain the perception of a smooth moving sound rendering without annoying discontinuities creating clicks in the 3D sound.

~~HRTFs~~HRIRs corresponding to specific azimuth and elevation can be synthesized by linearly combining a set of so-called Eigen-transfer functions (EFs) and a set of spatial characteristic functions (SCFs) for the relevant auditory space, as shown in Fig. ~~3~~1 herein, and as described in "An Implementation of Virtual Acoustic Space For Neurophysiological Studies of Directional Hearing" by Richard A. Reale, Jiashu Chen et al. in Virtual Auditory Space: Generation and Applications, edited by Simon Carlile (1996); and "A Spatial Feature Extraction and Regularization Model for the Head-Related Transfer Function" by Jiashu Chen et al. in J. Acoust. Soc. Am. 97 (1) (January 1995), the entirety of both of which are explicitly incorporated herein by reference.

In accordance with the principles of the present invention, spatial attributes extracted from the HRTFs are regularized before combination with the Eigen transfer function filters to provide a plurality of HRTFs with varying degrees of smoothness and generalization.

Fig. 1 shows an implementation of the regularization of a number N of SCF sample sets **202-206** in an otherwise conventional system as shown in Fig. 3.

In particular, a plurality N of Eigen filters **222-226** are associated with a corresponding plurality N of SCF samples **202-206**. A plurality N of regularizing models **212-216** act on the plurality N of SCF samples **202-206** before the SCF samples **202-206** are linearly combined with their corresponding Eigen filters **222-226**. Thus, in accordance with the principles of the present invention, SCF sample sets are regularized or smoothed before combination with their corresponding Eigen filters.

The particular level of smoothness desired can be controlled with a smoothness control to all regularizing models **212-216**, to allow the user to adjust a tradeoff between smoothness and localization of the sound image. The regularizing models **212-216** in the disclosed embodiment performs a so-called 'generalized spline model' function on the SCF sample sets **202-206**, such that smoothed continuous SCF sets are generated at combination points **230-234**, respectively. The degree of smoothing, or regularization, can be controlled by a lambda factor, with trade-offs of the smoothness of the SCF samples with their acuity.

The results of the combined Eigen filters **222-226** and corresponding regularized SCF sample sets **202-206/212-216** are summed in a summer **240**. The summed output from the summer **240** provides a single regularized HRTF (or HRIR) filter **250** through which the digital audio sound source **260** is passed, to provide an HRTF (or HRIR) filtered output **262**.

The HRTF filtering in a 3D sound system in accordance with the principles of the present invention may be performed either before or after other 3D sound processes, e.g., before or after an interaural delay is inserted into an audio signal. In the disclosed embodiment, the HRTF modeling process is performed after insertion of the interaural delay.

The regularizing models **212-216** are controlled by a desired location of the sound source, e.g., by varying a desired source elevation and/or azimuth.

Fig. 2 shows an exemplary process of providing the Eigen functions for the Eigen filters **222-226** and the SCF sample sets **202-206**, e.g., as shown in Fig. 1, to provide an HRTF model having varying degrees of smoothness and generalization in accordance with the principles of the present invention.

In particular, in step **102**, the ear canal impulse responses and free field response are measured from a microphone embedded in a

mannequin or human subject. The responses are measured with respect to a broadband stimulus sound source that is positioned at a distance about 1 meter or farther away from the microphone, and preferably moved in 5 to 15 degree intervals both in azimuth and elevation in a sphere.

In step **104**, the data measured in step **102** is used to derive the ~~HRTFs~~HRIRs using a discrete Fourier Transform (DFT) based method or other system identification method. Since the ~~HRTFs~~HRIRs are either in a frequency or time domain form, and since they vary with respect to their respective spatial location, ~~HRTFs~~HRIRs are generally considered as a multivariate function with frequency (or time) and spatial (azimuth and elevation) attributes.

In step **106**, an HRTF data covariance matrix is constructed either in the frequency domain or in the time domain. For instance, in the disclosed embodiment, a covariance data matrix of measured head-related impulse responses (HRIR) are measured.

In step **108**, an Eigen decomposition is performed on the data covariance matrix constructed in step **106**, to order the Eigen vectors according to their corresponding Eigen values. These Eigen vectors are a function of frequency only and are abbreviated herein as "EFs". Thus, the ~~HRTFs~~HRIRs are expressed as weighted combinations of a set of complex valued Eigen transfer functions (EFs). The EFs are an orthogonal set of frequency-dependent functions, and the weights applied to each EF are functions only of spatial location and are thus termed spatial characteristic functions (SCFs).

In step **110**, the principal Eigen vectors are determined. For instance, in the disclosed embodiment, an energy or power criteria may be used to select the N most significant Eigen vectors. These principal Eigen vectors form the basis for the Eigen filters **222-226** (Fig. 1).

In step **112**, all the measured ~~HRTFs~~HRIRs are back-projected to the principal Eigen vectors selected in step **110** to obtain N

sets of weights. These weight sets are viewed as discrete samples of N continuous functions. These functions are two dimensional with their arguments in azimuthal and elevation angles. They are termed spatial characteristic functions (SCFs). This process is called spatial feature extraction.

Each HRTF, either in its frequency or in its time domain form, can be re-synthesized by linearly combining the Eigen vectors and the SCFs. This linear combination is generally known as Karhunen-Loeve expansion.

Instead of directly using the derived SCFs as in conventional systems, e.g., as shown in Fig. 3, they are processed by a so-called "generalized spline model" in regularizing models **212-216** such that smoothed continuous SCF sets are generated at combinatorial points **230-234**. This process is referred to as spatial feature regularization. The degree of smoothing, or regularization, can be controlled by a smoothness control with a lambda factor, providing a trade-off between the smoothness of the SCF samples **202-206** and their acuity.

In step **114**, the measured HRIRs are back-projected to the principal Eigen vectors selected in step **110** to provide the spatial characteristic function (SCF) sample sets **202-206**.

Thus, in accordance with the principles of the present invention, SCF samples are regularized or smoothed before combination with a corresponding set of Eigen filters **222-226**, and recombined to form a new set of ~~HRTFs.~~HRIRs.

In accordance with the principles of the present invention, an improved set of ~~HRTFs~~HRIRs are created which, when used to generate moving sound, do not introduce discontinuities causing the annoying effects of clicking sound. Thus, with empirically selected lambda values, localization and smoothness can be traded off against one another to eliminate discontinuities in the ~~HRTFs.~~HRIRs.

While the invention has been described with reference to the exemplary embodiments thereof, those skilled in the art will be able to make various modifications to the described embodiments of the invention without departing from the true spirit and scope of the invention.

CLAIMS

What is claimed is:

1. A head-related transfer function model for use with 3D sound applications, comprising:

a plurality of Eigen filters;

a plurality of spatial characteristic functions are adapted to be respectively combined with said plurality of Eigen filters; and

a plurality of regularizing models adapted to regularize said plurality of spatial characteristic functions prior to said respective combination with said plurality of Eigen filters.

2. The head-related transfer function model for use with 3D sound applications according to claim 1, further comprising:

a summer ~~operably coupled to~~adapted to sum said plurality of combined Eigen filters combined with said plurality of regularized spatial characteristic functions to provide said head-related transfer function model.

3. The head-related transfer function model for use with 3D sound applications according to claim 1, wherein:

said plurality of regularizing models are each adapted to perform a generalized spline model.

4. The head-related transfer function model for use with 3D sound applications according to claim 1, further comprising:

a smoothness control ~~operably coupled~~in communication with said plurality of regularizing models to allow control of a trade-off between localization and smoothness of said head-related transfer function.

5. A head-related impulse response model for use with 3D sound applications, comprising:

a plurality of Eigen filters;

a plurality of spatial characteristic functions are adapted to be respectively combined with said plurality of Eigen filters; and

a plurality of regularizing models adapted to regularize said plurality of spatial characteristic functions prior to said respective combination with said plurality of Eigen filters.

6. The head-related impulse response model for use with 3D sound applications according to claim 5, further comprising:

a summer adapted to sum said plurality of combined Eigen filters combined with said plurality of regularized spatial characteristic functions to provide said head-related impulse response model.

7. The head-related impulse response model for use with 3D sound applications according to claim 5, wherein:

said plurality of regularizing models are each adapted to perform a generalized spline model.

8. The head-related transfer function model for use with 3D sound applications according to claim 5, further comprising:

a smoothness control in communication with said plurality of regularizing models to allow control of a trade-off between localization and smoothness of said head-related transfer function.

9. A method of determining spatial characteristic sets for use in a head-related transfer function model, comprising:

constructing a covariance data matrix of a plurality of measured head-related transfer functions;

performing an Eigen decomposition of said covariance data matrix to provide a plurality of Eigen vectors;

determining at least one principal Eigen vector from said plurality of Eigen vectors; and

back-projecting said measured head-related transfer functions~~back~~ to said at least one principal Eigen vector to create said spatial characteristic sets.

10. A method of determining spatial characteristic sets for use in a head-related impulse response model, comprising:

constructing a covariance data matrix of a plurality of measured head-related impulse responses;

performing an Eigen decomposition of said covariance data matrix to provide a plurality of Eigen vectors;

determining at least one principal Eigen vector from said plurality of Eigen vectors; and

back-projecting said measured head-related impulse responses to said at least one principal Eigen vector to create said spatial characteristic sets.

11. Apparatus for determining spatial characteristic sets for use in a head-related transfer function model, comprising:

means for constructing a covariance data matrix of a plurality of measured head-related transfer functions;

means for performing an Eigen decomposition of said covariance data matrix to provide a plurality of Eigen vectors;

means for determining at least one principal Eigen vector from said plurality of Eigen vectors; and

means for back-projecting said measured head-related transfer functions to said at least one principal Eigen vector to create said spatial characteristic sets.

12. Apparatus for determining spatial characteristic sets for use in a head-related impulse response model, comprising:

means for constructing a covariance data matrix of a plurality of measured head-related impulse responses;

means for performing an Eigen decomposition of said covariance data matrix to provide a plurality of Eigen vectors;

means for determining at least one principal Eigen vector from said plurality of Eigen vectors; and

means for back-projecting said measured head-related impulse responses to said at least one principal Eigen vector to create said spatial characteristic sets.

ABSTRACT

The present invention provides an improved HRTF modeling technique for synthesizing HRTFs with varying degrees of smoothness and generalization. A plurality N of spatial characteristic function sets are regularized or smoothed before combination with corresponding Eigen filter functions, and summed to provide an HRTF (or HRIR) filter having improved smoothness in a continuous auditory space. A trade-off is allowed between accuracy in localization and smoothness by controlling the smoothness level of the regularizing models with a lambda factor. Improved smoothness in the HRTF filter allows the perception by the listener of a smoothly moving sound rendering free of annoying discontinuities creating clicks in the 3D sound.